Classification of Wolf Rayet stars using Ensemble-based Machine Learning algorithms

Rajorshi Bhattacharya (UNM/NRAO)
Subhajit Kar (SNBNCBS)
Ylva Pihlstrom (UNM/NRAO)
Ramkrishna Das (SNBNCBS)
Megan Lewis (Leiden observatory, Leiden University)

Introduction

- Wolf-Rayet (WR) stars belong to the evolved class of Population I stellar objects, mainly arising from the evolutionary trajectory of massive O-type Main Sequence (MS) stars.

- From IR and optical surveys a significant population (670) of WR stars in the Milky Way (MW) has been detected, while simple population models predict that around 2000 WR stars may be present.

- WR stars often occur in clusters, making them suitable tracers of massive star formation regions in the MW and in Starburst Galaxies .

Credits: Moore et. al. (2000)

Motivation

- From IR and optical surveys, a significant population (670) of WR stars in the MW has been detected.
  - Simple population models predict that around 2000 WR stars may be present.

- Distinguishing WR stars from other stellar objects solely based on optical colors and magnitudes has been challenging because of strong interstellar dust extinction.

- Meanwhile, the IR bands suffer from much less obscuration than the optical bands, thus benefiting the detection of WR stellar candidates.

- With the advent of modern instruments such as JWST, Roman telescope, etc., we are bound to discover several thousands of such evolved stellar objects lying in the Local Group of galaxies.

- Selection criteria based on the IR (2MASS and GLIMPSE) colors and magnitudes have been successful in filtering non-WR from WR candidates (Mauerhan et.al. 2011).

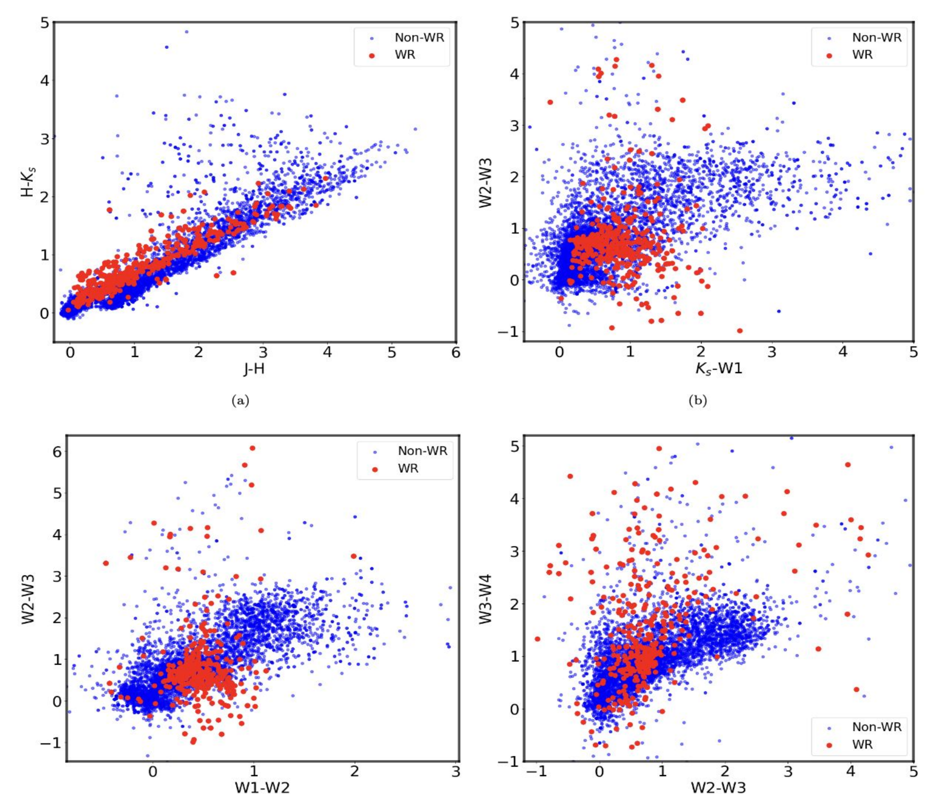
- Machine learning (ML) techniques are well-suited for handling large datasets and offer a powerful and flexible approach to stellar classification.

- Recent studies have identified several WR stars using the narrow-band IR (2MASS and GLIMPSE) magnitudes and colors using ML models (Morello et.al. 2018).

Selecting IR colors

- We use the 2MASS and WISE colors for our ML model
  - Allows us to get a larger data-set

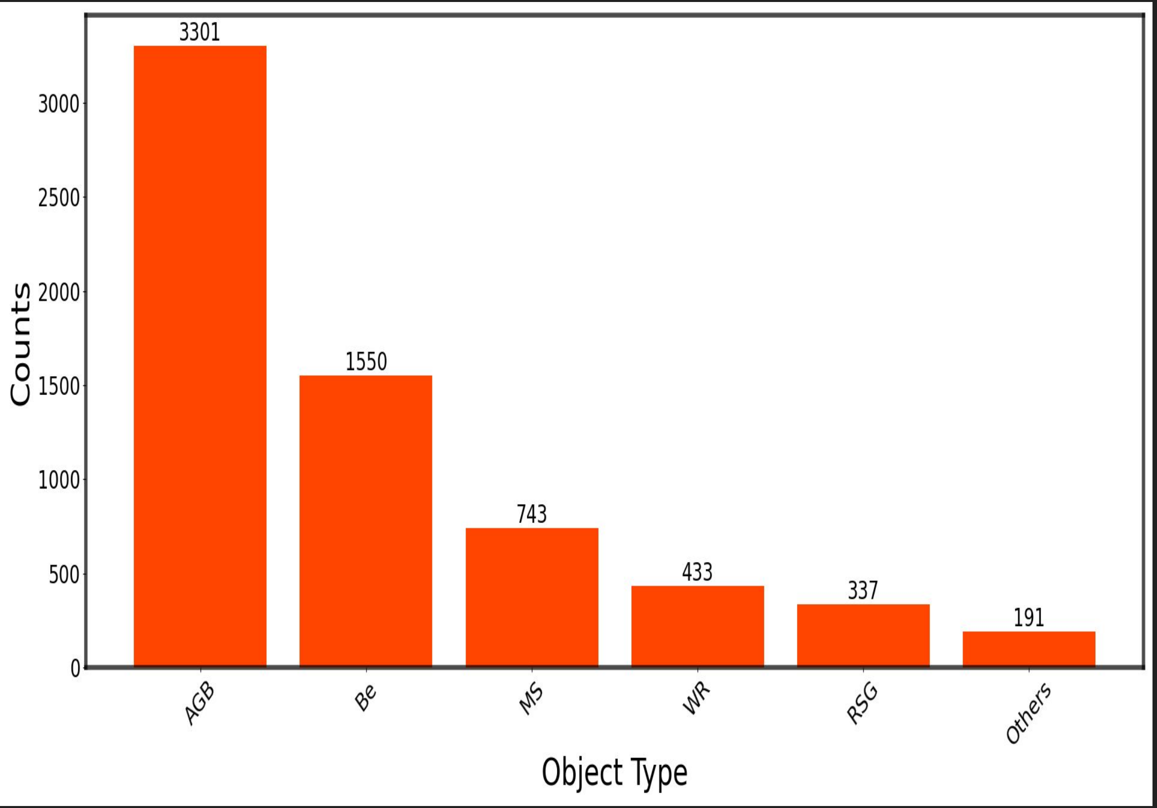
- We choose a large dataset (13,186) of stellar objects using Simbad which consists of WR stars, Main Sequence (MS), Asymptotic Giant Branch (AGB), Red Super Giant (RSG), Be-type stars etc.



Data sources

- We cleaned the dataset by choosing Galactic stellar objects with available apparent magnitudes in both 2MASS and WISE bands with good quality flags.

- Employed the color cuts from Mauerhan et.al.(2011) to remove extreme outliers.
    - Training and testing data: 6555 objects.



Kar and Bhattacharya et.al.(2024)

Feature selection

- Our ML model included features that consisted of IR colors ($J - H$, $H - K_s$, K_s -W1 ,W1− W2, W2 − W3, and W3 − W4) and positional attributes (RA and DEC).

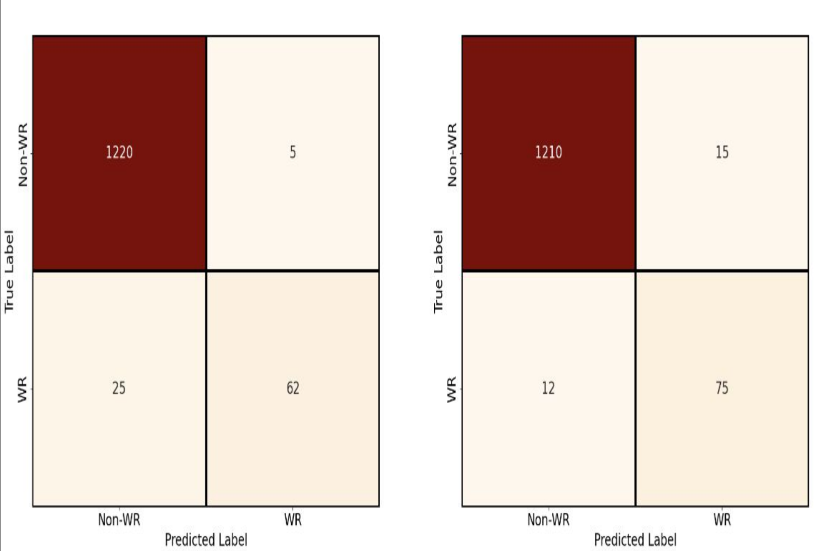
- For this work, we do not include the magnitudes, as neither distance estimates nor extinction values are known for the whole sample.

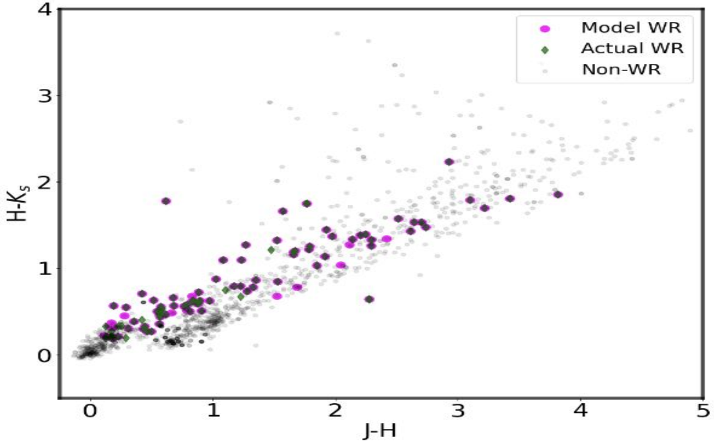
- Although IR colors can be affected by distance-dependent extinction, this effect is much smaller than the effect on magnitudes or even optical colors.

Results

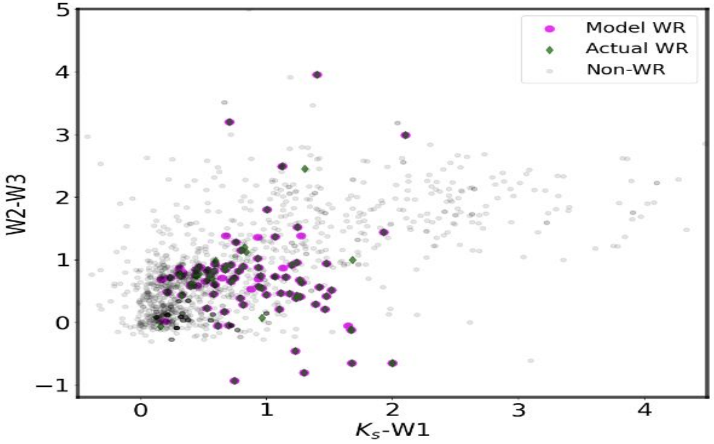
Table 3. Performance statistics of Object classifier models applied on the TsD-1.

Metric	XGB model	RF model
(1)	(2)	(3)
Accuracy	98	98
Precision	83	92
Recall	86	71
f1	84	80

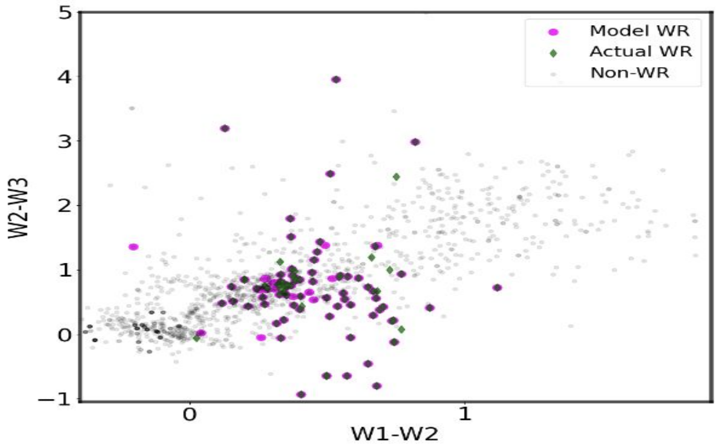


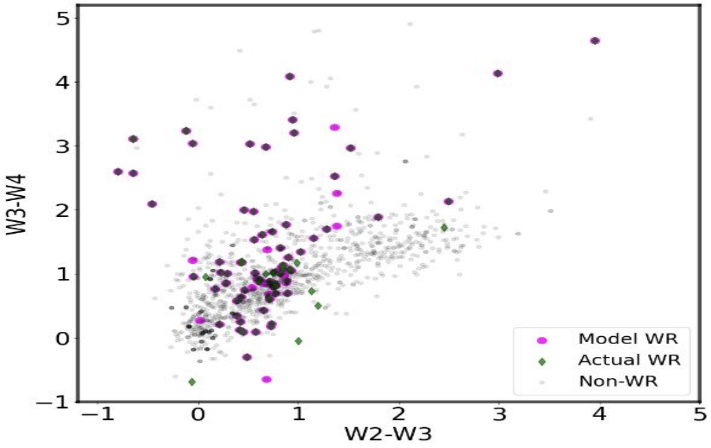

Model WR
Actual WR
Non-WR
H-K_s
J-H

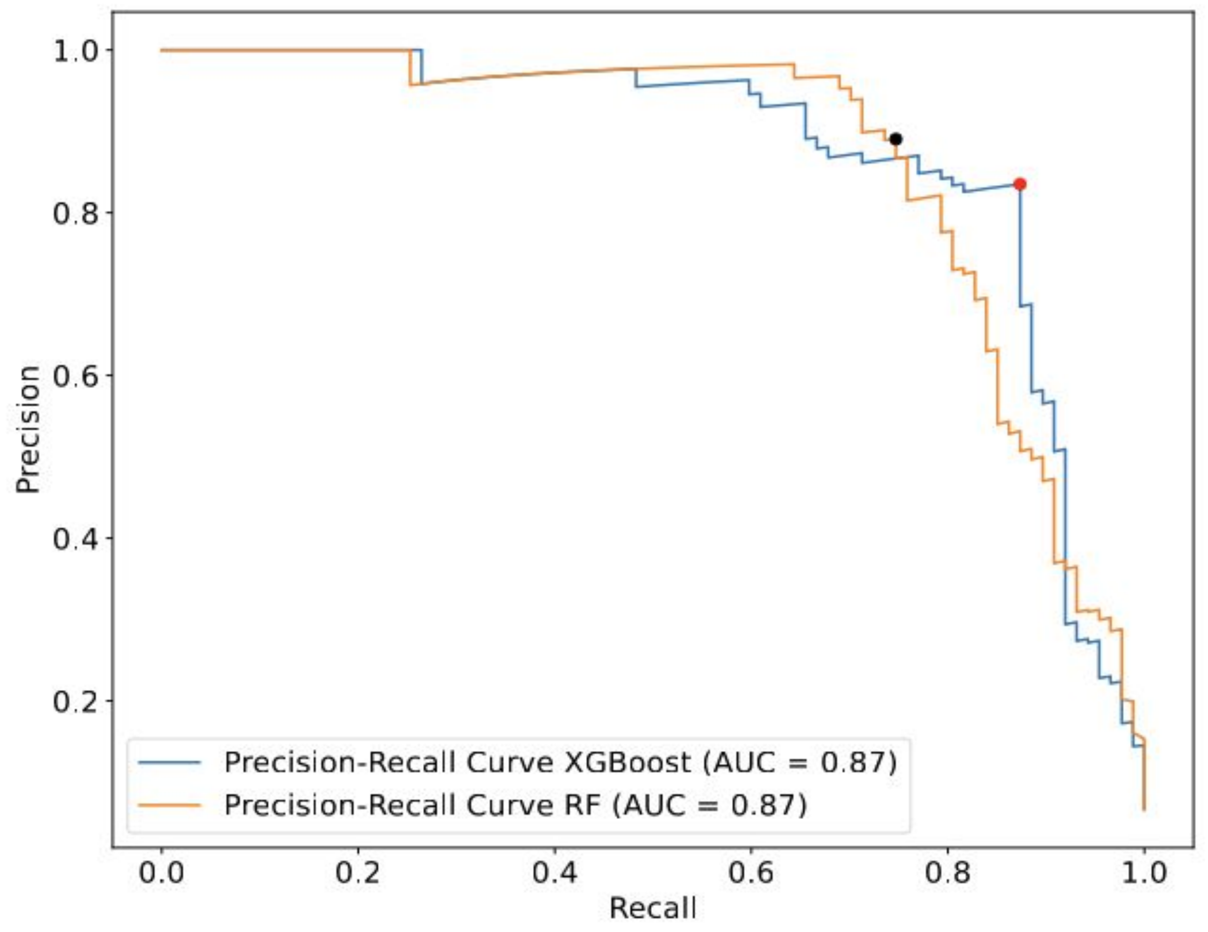
(a)


Model WR
Actual WR
Non-WR
W2-W3
K_s-W1

(b)


Model WR
Actual WR
Non-WR
W2-W3
W1-W2


Model WR
Actual WR
Non-WR
W3-W4
W2-W3

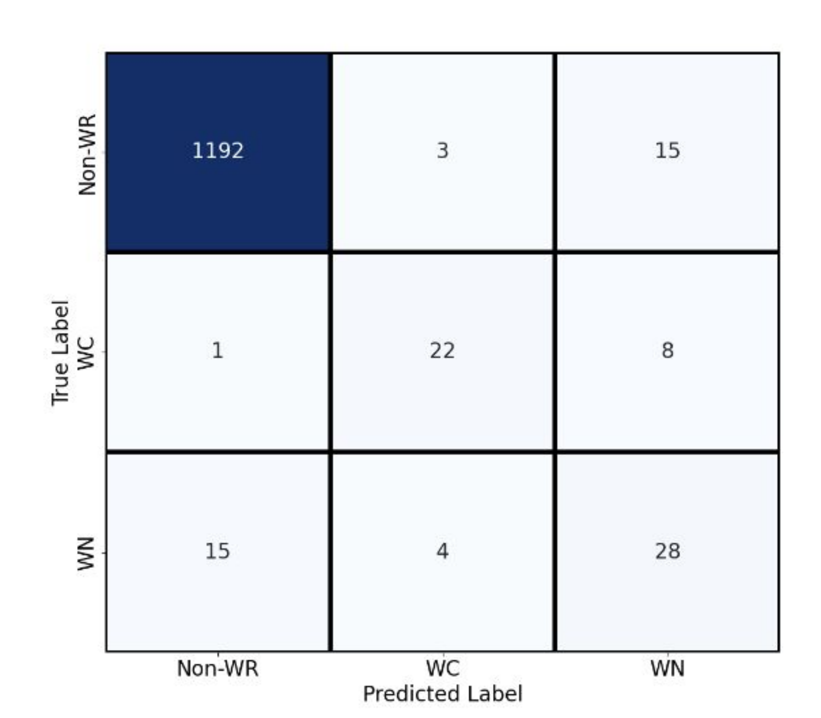


Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8
1.0
Precision-Recall Curve XGBoost (AUC = 0.87)
Precision-Recall Curve RF (AUC = 0.87)

Classifying WR stars based on their spectral types

Table 5. Performance statistics of WR subtype classifier model applied on the TsD-2.

Metric	WC-class	WN-class
(1)	(2)	(3)
Precision	76	55
Recall	71	60
f1	73	57



Discussion

- We find that our models can identify WR stars across the broad IR color space with enhanced accuracy than what was achieved from selecting WR stars from so-called sweet-spot in the color-color diagrams.
  - Additionally, both of our models (XGB and RF) perform much better than the earlier KNN (Morello et.al.(2018)) and SVM (Dorn-Wallenstein et.al.(2021))..

- We find a better f1-score for XGB compared to RF.
  - XGB has better recall than RF

- Both the classifiers predict that the major contaminants for WR sources in the NIR color space are the Be-type objects which was also reported in Faherty et.al.(2014).
  - In mid-IR O-rich AGBs are the main contaminant.

- We find that the subtype XGB-classifier model classifies both the WC and WN type objects from the non-WR sources with good accuracy rates.

- For WC-type stars, most of the TPs detected by the classifier are of WCL-type. TPs corresponding to the WN-type stars span across WNE to WNL-type.
    - The WNL-type objects dominate the list as they exhibit color-excess in the W3-W4 space due to free-free emission.

- WNL and WCL show similar color patterns leading to the observed misclassifications by the model.
    - Free-free emission or circumstellar dust formed in earlier stages of evolution (Crowther et.al.(2003)).

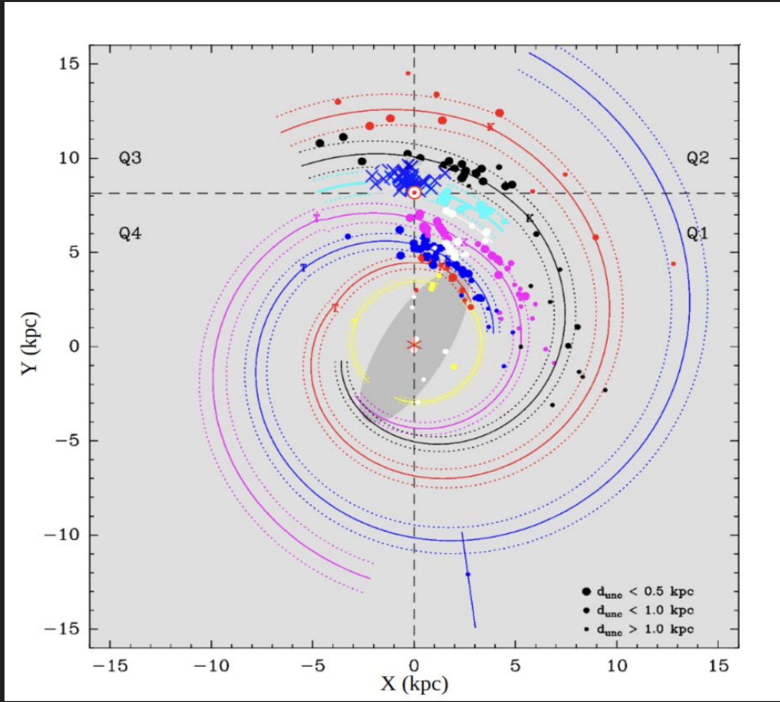
A new catalog of WR stars

- We apply our models on a large and different dataset from SIMBAD comprising stars with unknown object types.
- After cleaning and sorting the data in a similar manner we end up with 6457 objects.
- The model classifies 58 objects as WR stars .
- Subtype classifier model predict Subtypes for 10 of these sources.

- WR stars are known to be tracers of star formation.

- We use the Gaia parallaxes for our WR star candidates to see where they are in the MW.

- Most of our candidates lie on the local arm which is known to be an active massive SFR.
    - Might indicate that most of our candidates are actually WR stars.



Conclusion and Future Work

- We performed ML classification of WR stars from a large dataset (6555) of stellar objects of types ranging from MS to AGBs.
- We also developed a novel WR-subtype classifier using the XGB algorithm capable of distinguishing between WN and WC subtypes with exceptional accuracy.
- We also applied the object classifier model on an unlabelled dataset (comprised of 6457 stellar sources) and detected 58 new WR candidates.
- In future work, we plan to do a spectroscopic follow-up for our WR candidates and develop ML classifiers using the same algorithms to identify WR stars across metal-rich galaxies in the local group.